Encyclopedia of Research Design

Bootstrapping

Contributors: Christiana Hilmer
Editors: Neil J. Salkind
Book Title: Encyclopedia of Research Design
Chapter Title: "Bootstrapping"
Pub. Date: 2010
Access Date: May 05, 2014
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781412961271
Online ISBN: 9781412961288
DOI: http://dx.doi.org/10.4135/9781412961288.n34
Print pages: 102-105

http://dx.doi.org/10.4135/9781412961288.n34

The bootstrap is a computer-based statistical technique that is used to obtain measures of precision of parameter estimates. Although the technique is sufficiently general to be used in time-series analysis, permutation tests, cross-validation, nonlinear regression, and cluster analysis, its most common use is to compute standard errors and confidence intervals. Introduced by Bradley Efron in 1979, the procedure itself belongs in a broader class of estimators that use sampling techniques to create empirical distributions by resampling from the original data set. The goal of the procedure is to produce analytic expressions for estimators that are difficult to calculate mathematically. The name itself derives from the popular story in which Baron von Munchausen (after whom Munchausen syndrome is also named) was stuck at the bottom of a lake with no alternative but to grab his own bootstraps and pull himself to the surface. In a similar sense, when a closed-form mathematical solution is not easy to calculate, the researcher has no alternative but to “pull himself or herself up by the bootstraps” by employing such resampling techniques. This entry explores the basic principles and procedures of bootstrapping and examines its other applications and limitations.

Basic Principles and Estimation Procedures

The fundamental principle on which the procedure is based is the belief that under certain general conditions, the relationship between a bootstrapped estimator and a parameter estimate should be similar to the relationship between the parameter estimate and the unknown population parameter of interest. As a means of better understanding the origins of this belief, Peter Hall suggested a valuable visual: a nested Russian doll. According to Hall's thought experiment, a researcher is interested in **[p. 102 ↓]** determining the number of freckles present on the outermost doll. However, the researcher is not able to directly observe the outermost doll and instead can only directly observe the inner dolls, all of which resemble the outer doll, but because of their successively smaller size, each possesses successively fewer freckles. The question facing the researcher then is how to best use information from the observable inner dolls to draw conclusions about the likely number of freckles present on the outermost doll. To see how this works, assume for simplicity that the Russian doll set consists of three parts, the outermost doll and two inner dolls. In this case, the outermost doll can be thought of as the population, which is assumed to possess n_0 freckles; the second doll can be thought of as the original sample, which is assumed to possess n_1 freckles; and the third doll can be thought of as the bootstrap sample, which is assumed to possess n_2 freckles. A first guess in this situation might be to use the observed number of freckles on the second doll as the best estimate of the likely number of freckles on the outermost doll. Such an estimator will necessarily be biased, however, because the second doll is smaller than the outermost doll and necessarily possesses a smaller number of freckles. In other words, employing n_1 as an estimate of n_0 necessarily underestimates the true number of freckles on the outermost doll. This is where the bootstrapped estimator, n_2, reveals its true value. Because the third doll is smaller than the second doll by an amount similar to that by which the second doll is smaller than the outermost doll, the ratio of the number of freckles on the two inner dolls, $n_1 : n_2$, should be a close approximation of the ratio of the number of freckles on the second doll to number on the outer doll, n

0

: *n*
1

. This in a nutshell is the principle underlying the bootstrap procedure.

More formally, the nonparametric bootstrap derives from an empirical distribution function,
$\hat{F}$

, which is a random sample of size *n* from a probability distribution *F.* The estimator,
$\hat{\theta}$

, of the population parameter θ is defined as some function of the random sample (X_1, X_2, … X_n). The objective of the bootstrap is to assess the accuracy of the estimator,
$\hat{\theta}$

. The bootstrap principle described above states that the relationship between
$\hat{\theta}$

and θ should be mimicked by that between θ^b and
θ^b

, where θ^b is the bootstrap estimator from bootstrap samples. In practice, bootstrap samples are obtained by a Monte Carlo procedure to draw (with replacement) multiple random samples of size *n* from the initial sample data set, calculating the parameter of interest for the sample drawn, say θ^b, and repeating the process *k* times. Hence, the bootstrap technique allows researchers to generate an estimated sampling distribution in cases in which they have access to only a single sample rather than the entire population. A minimum value for *k* is typically assumed to be 100 and can be as many as 10,000, depending on the application.

Peter Bickel and David Freedman defined the following three necessary conditions if the bootstrap is to provide consistent estimates of the asymptotic distribution of a parameter: (1) The statistic being bootstrapped must converge weakly to an asymptotic distribution whenever the data-generating distribution is in a neighborhood of the truth, or in other words, the convergence still occurs if the truth is allowed to change within the neighborhood as the sample size grows. (2) The convergence to the asymptotic distribution must be uniform in that neighborhood. (3) The asymptotic distribution must depend on the data-generating process in a continuous way. If all three conditions hold, then the bootstrap should provide reliable estimates in many different applications.

As a concrete example, assume that we wish to obtain the standard error of the median value for a sample of 30 incomes. The researcher needs to create 100 bootstrap samples because this is the generally agreed on number of replications needed to compute a standard error. The easiest way to sample with replacement is to take the one data set and copy it 500 times for 100 bootstrap samples in order to guarantee that each observation has an equal likelihood of being chosen in each bootstrap sample. The researcher then assigns random numbers to each of the 15,000 observations (500 * 30) and sorts each observation by its random number assignment from lowest to highest. The next step is to make 100 bootstrap samples of 30 observations each and disregard the other 12,000 observations. After the 100 bootstrap samples have been made, the median is calculated from each of the samples, and the bootstrap estimate of the standard error is just the standard deviation of the 100 bootstrapped medians. Although this procedure may seem complicated, it **[p. 103 ↓]** is actually relatively easy to write a bootstrapping program with the use of almost any modern statistical program, and in fact, many statistical programs include a bootstrap command.

Besides generating standard error estimates, the bootstrap is commonly used to directly estimate confidence intervals in cases in which they would otherwise be difficult to produce. Although a number of different bootstrapping approaches exist for computing confidence intervals, the following discussion focuses on two of the most popular. The first, called *the percentile method*, is straightforward and easy to implement. For illustration purposes, assume that the researcher wishes to obtain a 90% confidence interval. To do so, the researcher would (a) start by obtaining 1,000 bootstrap samples and the resulting 1,000 bootstrap estimates,

$$[\hat{\theta} - T_{.95}^{boot}\ s.e.(\hat{\theta}), \hat{\theta} - T_{.05}^{boot} s.e.(\hat{\theta})].$$

, and (b) order the 1,000 observed estimates from the smallest to the largest. The 90% confidence interval would then consist of the specific value bootstrap estimates falling at the 5th and the 95th percentiles of the sorted distribution. This method typically works well for large sample sizes because the bootstrap mimics the sampling distribution, but it does not work well for small sample size. If the number of observations in the sample is small, Bradley Efron and Robert Tibshirani have suggested using a bias correction factor.

The second approach, called the *bootstrap* t *confidence interval*, is more complicated than the percentile method, but it is also more accurate. To understand this method, it is useful to review a standard confidence interval, which is defined as

$$[\hat{\theta} - t_{\alpha/2,df} s.e.(\hat{\theta}), \hat{\theta} + t_{\alpha/2,df} s.e.(\hat{\theta})]$$

, where

$$\hat{\theta}$$

is the estimate, *t*
#/2,df

is the critical value from the *t*-table with *df* degrees of freedom for a (1–$_{\alpha)}$ confidence interval, and *s.e.*

$$(\hat{\theta})$$

is the standard error of the estimate. The idea behind the bootstrap *t* interval is that the critical value is found through bootstrapping instead of simply reading the value contained in a published table. Specifically, the bootstrap *t* is defined as T^{boot} =

$$(\hat{\theta}^{boot} - \hat{\theta})$$

$/S^{boot}$, where

$$\hat{\theta}^{boot}$$

is the estimate of θ from a bootstrap sample and S^{boot} is an estimate of the standard deviation of θ from the bootstrap sample. The *k* values of T^{boot} are then ordered from lowest to highest, and then, for a 90% confidence interval, the value at the 5th percentile is the lower critical value and the value at the 95th percentile is the higher critical value. Thus the bootstrapped *t* interval is

$$[\hat{\theta} - T_{.95}^{boot}\ s.e.(\hat{\theta}), \hat{\theta} - T_{.05}^{boot} s.e.(\hat{\theta})].$$

Michael Chernick has pointed out that the biggest drawback of this method is that it is not always obvious how to compute the standard errors, S^{boot} and *s.e.*

$$(\hat{\theta})$$

.

Other Applications

In addition to calculating such measures of precision, the bootstrap procedure has gained favor for a number of other applications. For one, the bootstrap is now popular as a method for performing bias reduction. Bias reduction can be explained as follows. The bias of an estimator is the difference between the expected value of an estimator, E(

$\hat{\theta}$

), and the true value of the parameter, θ, or E(

$\hat{\theta} - \theta$

–θ). If an estimator is biased, then this value is nonzero, and the estimator is wrong on average. In the case of such a biased estimator, the bootstrap principle is employed such that the bias is estimated by taking the average of the difference between the bootstrap estimate,

$\hat{\theta}^b$

^{b}, and the estimate from the initial sample,

$\hat{\theta}$

over the *k* different bootstrap estimates. Efron defined the bias of the bootstrap as E(

$\hat{\theta}$

–

$\hat{\theta}$

^{b}) and suggested reducing the bias of the original estimator,

$\hat{\theta}$

, by adding estimated bias. This technique produces an estimator that is close to unbiased.

Recently, the bootstrap has also become popular in different types of regression analysis, including linear regression, nonlinear regression, time-series analysis, and forecasting. With linear regression, the researcher can either fit the residuals from the fitted model, or the vector of the dependent and independent variables can be bootstrapped. If the error terms are not normal and the sample size is small, then the researcher is able to obtain bootstrapped confidence intervals, like the one described above, instead of relying on asymptotic theory that likely does not apply. In nonlinear regression analysis, the bootstrap is a very useful tool because there is no need to differentiate and an analytic expression is not necessary.

[p. 104 ↓]

Limitations

Although the above discussion has highlighted that the bootstrap technique is potentially valuable in a number of situations, it should be noted that it is not the ideal solution to every statistical problem. One problem would occur in cases in which parameters are constrained to be on a boundary of the parameter space (such as when a priori theoretical restrictions require a certain estimated parameter to be of a specific sign). Common examples of such restrictions include traditional demand analysis in which the income effect for a normal good is constrained to be positive and the own-price effect is constrained to be negative, cost function analysis in which curvature constraints imply that second-order price terms satisfy concavity conditions, and time-series models for conditional heteroskedasticity in which the same parameters are constrained to be nonnegative. Such cases are potentially problematic for the researcher because standard error estimates and confidence bounds are difficult to compute using classical statistical inference, and therefore the bootstrap would be a natural choice. Don Andrews has demonstrated, however, that this procedure is not asymptotically correct to the first order when parameters are on a boundary. This is because the bootstrap puts too much mass below the cutoff point for the parameter and therefore does a poor job of mimicking the true population distribution. Other circumstances in which the bootstrap fails include an extremely small sample size, its use with matching estimators to evaluate programs, and distributions with long tails.

ChristianaHilmer

http://dx.doi.org/10.4135/9781412961288.n34
See also

- Bias
- Confidence Intervals
- Distribution
- Jackknife
- Central Tendency, Measures of
- Median
- Random Sampling
- Sampling
- Standard Deviation
- Standard Error of Estimate
- Statistic
- Student's *t* Test
- Unbiased Estimator
- Variability, Measure of

Further Readings

Abadie, A., &Imbens, G. W. (2006). On the failure of the bootstrap for matching estimators (NBER Working Paper 0325). Cambridge, MA: National Bureau of Economic Research.

Andrews, D. W. K. Estimation when a parameter is on a boundary. Econometrica, (1999). vol. 67(6), pp. 1341–1383.

Andrews, D. W. K. A three-step method for choosing the number of bootstrap repetitions. Econometrica, (2003). vol. 68(1), pp. 23–51.

Arthreya, K. B. Bootstrap of the mean in the infinite variance case. Annals of Statistics, (1985). vol. 15, pp. 724–731.

Bickel, P. J., & and Freeman, D. Some asymptotic theory for the bootstrap. Annals of Statistics, (1981). vol. 9, pp. 1196–1217.

Chernick, M. R. (1999). Bootstrap methods: A practitioner's guide. New York: Wiley.

Efron, B. Bootstrap methods: Another look at the jackknife. Annals of Statistics, (1979). vol. 7, pp. 1–26.

Efron, B. Estimating the error rate of a perdiction rule: Improvements on cross-validation. Journal of the American Statistical Association, (1983). vol. 78, pp. 316–331.

Efron, B. Better bootstrap confidence intervals, Journal of the American Statistical Association, (1987). vol. 82, pp. 171–200.

Efron, B., &Tibshirani, R. (1986). An introduction to the bootstrap. New York: Chapman & Hall.

Hall, P. (1992). The bootstrap and Edgeworth expansion. New York: Springer-Verlag.